Australian Government
Department of Defence
Defence Science and Technology Organisation

Change Detection in Rough Time Series

Lewis Warren

National Security and ISR Division
Defence Science and Technology Organisation

DSTO-TR-3023

ABSTRACT

A discrete time series may have high noise levels resulting in a rough or jagged distribution that can present significant challenges to conventional statistical tracking techniques. To address this problem the proposed method applies hybrid fuzzy statistical techniques to series granules instead of to individual measures. Three examples demonstrated the robust nature of the proposed fuzzy tracking signal that leads to a minimal number of false alarms caused by isolated spikes. These examples demonstrated the effectiveness of this tracking signal for promptly identifying significant pattern changes in rough time series as can be encountered in data sets used for various types of Defence decision making.

RELEASE LIMITATION

Approved for public release

Published by

National Security and ISR Division
DSTO Defence Science and Technology Organisation
PO Box 1500
Edinburgh South Australia 5111 Australia

Telephone: 1300 333 362
Fax: (08) 7389 6567

AR-016-091
September 2014

APPROVED FOR PUBLIC RELEASE

Change Detection in Rough Time Series

Executive Summary

This report presents a procedure for tracking a rough series of data with the objective of promptly detecting an unusual sequence of values in the series that indicates some sort of significant change. A rough series of data is defined as a jagged pattern with discontinuities and a relatively high noise level. Such a series may be sparse and result from sporadic event or measurement occurrences, and also may consist of many low or nil values. The data may represent approximate evaluations or inaccurate sensor data, subjective ratings of vague variables, imperfect intelligence reports, algorithmic derived measures indicating degrees of suspected network intrusions, or simply be data degraded by some form of interference.

This new tracking signal being proposed is based on a hybrid methodology with combined statistical and fuzzy set aspects. The proposed hybrid method is also based on information granulation which is simply a partitioning of information elements into groups. Three case studies were examined to demonstrate the robust nature of the proposed tracking signal and its ability to promptly detect anomalies. The three tracking examples are: detecting significant changes in a threat indicator derived from automated newsfeeds about North Korea; insurgent activity in Iraq; and in the rainfall pattern in the Murray-Darling Basin over the last 107 years. The results for the Murray-Darling Basin rainfall data also enabled some new conclusions to be inferred that could not be found in existing expert statistical analyses of that data set. Moreover, the examples demonstrate the effectiveness of this tracking signal for promptly identifying significant pattern changes in rough time series as can be encountered in data sets used for various types of Defence decision making.

This page is intentionally blank

Author

Lewis Warren
National Security and ISR Division

Lewis Warren has a background in Operations Research consulting, industrial simulation, mathematical methods for decision analysis, and performance evaluation of complex systems. His educational qualifications are: B. Industrial Eng.(Hons) from University of Melbourne, Master of Engineering(Systems) from RMIT University, and a PhD in Complex Systems Modelling from Swinburne University of Technology. In recent years his research has focused on soft OR techniques and methods for the representation and management of hybrid forms of uncertainty in decision support models. While at DSTO he has worked on applications for modelling strategic decisions, intelligence analysis, and decision support systems for full spectrum military operations planning and evaluation.

This page is intentionally blank

Contents

This page is intentionally blank

1. Introduction

This report presents a procedure for tracking a rough series of data with the objective of promptly detecting an unusual sequence of values in the series that indicates some sort of significant change. Generally speaking, this report will assume a rough time series to have the following characteristics which can be noted in the demonstration examples shown in Figures 3, 6, and 9:

- Data elements may not be accurate, or be qualitative as with subjective ratings.
- High random noise may exist due to measurement or various other causes.
- Large fluctuations of values around an overall trend line.
- Relatively low autocorrelation between successive values.
- Possible null values either as true values or from missing data.

This type of data may originate from many sources such as: imperfect intelligence reports, human estimates of qualitative variables, algorithmic derived measures indicating degrees of suspected network intrusions, inaccurate sensor data, indirect estimates from covariates of a target variable, or simply be data degraded by some form of interference. The target problem of this report is a special case of the general change-point detection problem in time series tracking. Various methods have been proposed to address this problem most of which assume that a considerable amount of data is available. Detecting unusual patterns in stock market indicators and credit card fraud detection have been fertile application areas. However, the rather sparse and jagged nature of rough series limits the application of many analytical methods which depend upon rich data streams.

To address this problem, a hybrid methodology is proposed with combined statistical and fuzzy set aspects. The proposed method is based upon information granulation which is simply a partitioning of information elements into groups. Information granulation has been applied across many types of modelling, using both formal and informal methods. Zadeh [1] has pointed out how information granulation is ubiquitous in human actions, cognition, and reasoning. Granules may be crisp or fuzzy, and fuzzy sets are themselves granules as groupings of information elements. The primary reason for adopting information granulation in the following procedure is to capture some of the uncertainty existing in the raw data elements.

The tracking problem we are considering here is to determine with minimal false alarms, when a notable change occurs in an unfolding data series. The purpose of applying an automated anomaly detection algorithm may be to authenticate, or reinforce, any detection that occurs from a manual inspection of the measure sequences. This is important for rough series since their inherent nature may disguise pattern changes in the unfolding data. Alternatively, an automated alert could prompt a visual inspection for confirming a pattern change. Three example applications demonstrate the behaviour of the proposed tracking signal. The first, of two Defence related applications, is to track a threat indicator for North Korea developed from automated analysis of newsfeeds. The second application is to detect changes in insurgent activity in Iraq. The third application is to detect anomalies in historical Murray-Darling Basin rainfall data. These applications demonstrate the robust nature and effectiveness of the proposed tracking signal for detecting anomalies in rough series.

2. Tracking Temporal Data

2.1 Overview of Tracking

Tracking of temporal data series is commonly applied in the commercial world for predictive purposes such as predicting future product demand so that optimal inventory levels can be established. However, there are many other reasons for tracking time series of unfolding temporal data, and another objective is to detect some kind of pattern change or deviation. A sample of tracking applications follows.

Process Quality Control: [2][3][4][5]
To ensure that a production process achieves desired product quality by application of statistical cumulative sum of error techniques (CUSUM) and checking against statistical control limits for a specified degree of confidence.

Forecasting: [6][7][8]
To monitor sales forecasting for production planning purposes and detect when forecasts are significantly different to actual sales so that production levels would be inappropriate.

Telecom Network Management:
To monitor traffic flow so that optimal ways of achieving desired service levels can be dynamically maintained [9][10], or to detect unusual patterns indicating illegal intrusions [11].

Mobile Phone Use or Credit Card Fraud Detection: [12]
To detect unusual activity on an account record that indicates an illegal user.

Financial Indicator Turning Points: [13]
To detect turning points so that Buy/Sell profits can be maximised.

Sequential Patterns in Medical Data: [14]
To discover common sequences between medical data sets, and to determine when such sequences occur for the identification of medical conditions.

Across all these applications two types of procedures may be applied: static analysis of data sets, or monitoring of dynamic series. Static analysis is what many data mining or knowledge discovery techniques perform, while monitoring of unfolding dynamic series is often referred to as tracking. Hybrid algorithms may also be applied which then contain both static analysis and dynamic monitoring mechanisms. The use of training data sets for model specification followed by the application of adaptive learning mechanisms is also a frequently encountered hybrid process. However, in this report the focus is only on the monitoring of dynamic series.

2.2 Tracking to Detect Estimator Bias

A large variety of algorithms have been proposed for tracking temporal data for forecasting purposes [8], or detecting when abrupt changes have occurred [15]. But many tracking algorithms assume that a large amount of data is available to establish behavioural patterns as do some data mining techniques. These approaches are fundamentally different to the monitoring procedure proposed in this report, the key difference being that the proposed method is designed for sparse data sets, which may also be of qualitative or subjective measures. In general, the basic rationale for tracking a time series is that each new input is used to update an estimator for the series. The estimator is most commonly a predictive value covering some time into the future (one step or more). However, the estimator may also be a current average, or some other measure relating to the current state. This is the type of estimator that is used in the proposed procedure. Running error values are then determined from the deviation between the estimator and what actually occurs. A tracking signal is then computed from these errors.

An early tracking signal proposed by Brown [6,7] was based on the running sum of errors as used in CUSUM statistical methods for process control. That method also assumed that a mean absolute deviation (MAD) could be used to approximate the standard deviation of errors (σ_e) as 1.2 MAD, to establish the control limits. But several authors noted that the distribution of errors is often not Normal so this approximation is often inaccurate. Also, in 1964 Trigg [16] noted that a tracking signal based on the sum of estimate errors was too sensitive and not effective in identifying sharp changes. To address this problem Trigg proposed comparing the average error, instead of the sum of errors, to the control limits which were still based on the standard deviation of errors (σ_e) as approximated by 1.2 MAD.

The method in this paper establishes the tracking signal control limits by assuming the distribution of *average* errors will be Normal, rather than the distribution of the errors themselves. This assumption is based on the Central Limit Theorem as stated in [17]:

> …the sample mean $\bar{X}$, of a large random sample from a population with mean μ_X, and standard deviation σ_X, possesses a sampling distribution that is approximately Normal regardless of the distribution of the population from which the sample is obtained. The larger the sample size, the better will be the Normal approximation to the sampling distribution of $\bar{X}$.

And as with many statistical tracking methods, the null hypothesis is that the *average* error is zero when the system is stationary, with no unusual patterns. Then, for a Normal distribution, 95% of the average errors (which are called here Bias) should be within ± 2 $\sigma_{\bar{e}}$, where $\sigma_{\bar{e}}$ is the standard deviation of the average errors. When the average error is outside these limits we can say with 95% confidence that is not a random fluctuation, and is due to some significant change in the series. This is the rationale for selecting a control limit of 2, although the limit is not rigid for such a rough series, and any value beyond 1.5 indicates with increasing confidence that the average error is not due to a random fluctuation. Strictly speaking, when applying statistical methods to small samples of less than 28 data points, the Student-t distribution should be applied rather than the Normal

distribution. However, because the following procedure is fundamentally qualitative, the small deviation in the degree of confidence at 2 sigma limits from that of the Student-t distribution is not significant for deciding if some significant change has occurred.

3. Proposed Hybrid Tracking Method

The steps in the proposed procedure for detecting pattern changes in rough time series will now be described. This procedure is a hybrid combination of fuzzy set theory and statistical estimation theory.

3.1 Input Data Pre-processing

Periodic tracking at regular intervals may require data in each intervening period to be pre-processsed into a new input measure due to discontinuities in the series. Two possible methods are presented below: the first method applies when only a single event (or no event) can occur in the base time interval, and the other more general case applies where multiple measures (or none at all) can occur in the base time interval.

3.1.1 Pre-processing of single data elements in a period

A discontinuous series may exist when there is a series of periods with no data. One pre-processing approach is first to sum the last 'k' periods values (say three) to decrease the discontinuity of the series. Another approach is simply to update estimates only when input data actually occurs. This approach would in general be better when there is missing data, rather than data with null values, because summing a number of periods effectively assumes any missing data has null values. For demonstration purposes, the following examples will aggregate data over the last three time periods as running sums of three. The main purpose for this type of pre-processing is to attenuate the effect on the tracking signal of null values in a period.

After any pre-processing that is required, the next step is to transform these aggregates (sums of three) into a fuzzy granule for estimation and tracking purposes. The key to the effectiveness of the proposed method is the use of fuzzy granulation which enables some of the rough data characteristics to be captured. This then enables robust signal behaviour to be achieved, without the over-sensitivity that frequently results when highly responsive tracking parameters are applied (as with a large smoothing constant in exponential smoothing for example).

3.1.2 Pre-processing multiple data elements in a period

When multiple data elements can occur in a single base period, some technique must first be applied to reduce these to a single summary measure to be used in the proposed procedure. Although additive averaging is a common approach to summarise this type of input data, such compensatory aggregation may be misleading, or else tend to hide some valuable information on data polarisation. If this kind of information is significant in the data set, a non-additive aggregation technique may be preferable for integrating the data elements.

3.2 Forming the Input Fuzzy Granule

Following any pre-processing, the input fuzzy granule for the most recent updating period is created. To do this, initially rank order the last three updating period measures, those measures perhaps already pre-processed. Consider the three ordered measures {a,b,c} as a triangular possibility distribution (TPD) as the input for the current updating period. The modal value 'b' then represents the most possible or most feasible value, and the first and last values of the triplet represent the boundary measures for the input data granule. Notably, the modal value is not necessarily midway between the two boundary values.

3.3 Updating the Fuzzy Average Granule

Next, the input data granule (TPD) is first-order exponentially smoothed to estimate the non-stationary fuzzy average of the data. This smoothing technique decreases the weights of historical data exponentially so is biased towards more recent updated fuzzy estimates. The sensitivity of the exponential smoothing technique is determined by the smoothing constant (α), and an equivalent moving average period (N) can be determined [18] by the relationship, $\alpha = 2/(1+N)$. For $\alpha = 0.4$ in the following examples, a similar moving average period is then 4 time periods. A description of the fuzzification of the standard exponential smoothing technique has been provided by Kaufmann and Gupta in Chapter 15 of [19].

3.4 A Fuzzy Enhanced Tracking Signal (FETS)

A fuzzified tracking signal is next calculated from the deviation of the average granule estimate from the current input data granule. The modal value (i.e. the most possible value) of the fuzzy tracking signal TPD is then used to determine if a significant pattern change in the series has occurred.

Exponential smoothing, or averaging, is frequently applied in techniques to predict future values. Tracking signals are also used to detect the persistence of deviations that indicate when an inappropriate model, or parameter, may require some adaptive correction to yield better forecasts. Because the purpose of applying exponential smoothing in this paper is to promptly detect when a significant change has occurred in a non-stationary series, as opposed to making forecasts, it is more important in our case to modify the tracking signal in order to minimise the *delay* in detecting change, while at the same time minimising false alarms. This is the rationale for selecting the smoothing constant values in the following method.

Among the different approaches that have been proposed for determining control statistics to track data, many apply conventional statistical control chart procedures [2][3][20], while others use log transforms [4] to address the fact that the Normal distribution assumption for the error distribution is often invalid. As previously described, one tracking signal [6] that has been widely used is the ratio of the running sum of errors over 1.2 times the Mean Absolute Deviation (MAD), as the average of unsigned error magnitudes. Model bias is then indicated when this ratio is 'large' compared to some control limits. However, without assuming a statistical error distribution it is difficult to say what the control limit values should be, and frequently they are set quite arbitrarily. For example, it has been

suggested [5] that such a limit for industrial inventory control applications should be set between –3 and +8.

Toelle [21,22] has proposed the Enhanced Tracking Signal (ETS) which is based on control limits calculated from the standard deviation of *average* errors. In fact, the ETS of Toelle is an extension of one proposed by Trigg [16] about 50 years ago, by using control limits based on the standard deviation of average errors rather than the MAD. After comparative tests with other methods on 191 standard industrial data sets, Toelle concluded [21] that the ETS was usually very effective for detecting change in many types of time series. Toelle's proposed ETS has been modified and extended here in the following ways to make it more robust and responsive for tracking rough time series.

- *Use of fuzzy estimates*
  The input data and all estimates are fuzzified as triangular possibility distributions {a,b,c} to reduce the effect of large transients or spikes. Only the modal value 'b' of the updated FETS is checked against the control limits. This contributes to its robustness by eliminating alerts generated by the 'a' and 'c' least possible values of the FETS
- *Different averaging technique*
  As well as for updating the series average, the average error (or Bias) is also estimated by exponential smoothing. This replaces the long-term arithmetic average as used in Toelle's method.
- *Two different averaging periods for updating estimates*
  The average error itself is estimated over a short time period ($\alpha = 0.4$) to discount older data. On the other hand, the standard deviation of the average error is estimated over the whole series length to capture all existing variability.

The TPD of the standard deviation of *average* errors ($\tilde{\sigma}_{\bar{e}}$) is estimated as follows, where $\tilde{x}_i$ are the input data granules, Average is the series fuzzy exponential average TPD, and 'n' is the time period count from start of series.

$$\tilde{\sigma}_{\bar{e}} = \sqrt{\frac{\tilde{\sigma}_e^2}{n}}$$

$$\text{and} \quad \tilde{\sigma}_e^2 = \frac{\sum_{i=1}^{n} (\text{ Average} - \tilde{x}_i)^2}{(n-1)}$$

$$\text{Thus,} \quad \tilde{\sigma}_{\bar{e}} = \sqrt{\frac{\sum_{i=1}^{n} (\text{ Average} - \tilde{x}_i)^2}{n(n-1)}}$$

With the above extensions to the ETS of Toelle, the updating steps in the proposed procedure using the input TPD granules are as follows. In these equations the following fuzzy operators are used as described by Kaufmann and Gupta [19]:

Fuzzy addition = $\oplus$
Fuzzy subtraction = $\varnothing$
Fuzzy multiplication = $\otimes$
Fuzzy division = (:)

1. New Average TPD = α Input TPD $\oplus$ (1 - α) Old Average TPD
   (This is the fuzzified standard exponential formula where α is the crisp smoothing constant.)

2. Calculate Error TPD at each update time:
   Error TPD = Old Average TPD $\varnothing$ Input TPD

3. Update Smoothed Average Error (New Bias) Estimate:
   New Bias TPD = α Error TPD $\oplus$ (1 - α) Old Bias TPD

4. Compute the TPD of (Error TPD) ^{2}

5. Update Running Sum of Error^2 : $\tilde{S} = \Sigma$ (Error TPD)^2

6. Compute Standard Deviation of the Average Error (Bias) :
$$\tilde{\sigma}_{\bar{e}} = \sqrt{\frac{\tilde{S}}{n(n-1)}}$$

7. Enhanced Tracking Signal TPD:
$$\text{FETS }\{a,b,c\} = \frac{\text{New Bias TPD}}{\tilde{\sigma}_{\bar{e}}} \text{ (by fuzzy division (:))}$$

Figures 1 and 2 illustrate the above procedure for calculating the FETS and determining when a significant change has occurred. The general mathematical application of fuzzy operators on triangular fuzzy numbers can be found in Appendix A as found in Dubois and Prade [23]. An Excel program has also been developed to implement the method.

3.5 Interpreting the Tracking Signal

It is assumed that the distribution of the *average* errors is Normal, whereas the error distribution itself need not be, as per the Central Limit Theorem. Then, when the magnitude of the FETS is beyond 2 a statistically significant bias can be said to exist at about the 95% confidence level. Moreover, values >1.5 indicate a bias with a high degree of confidence. The exact (or inexact) nature of the problem and data would actually determine what are suitable period lengths upon which to base the respective error estimates, especially if any periodicity is present. In the following examples, a smoothing constant of 0.4 imputes the short term to be 4 periods, while the running history at any point in time determines the longer period for standard deviation estimates. In other cases different values could be appropriate for the short and long periods. A pattern change is then indicated by a sequence of control limit transgressions by the FETS mode, and two in a row generates an alert in the following examples. This fuzzy statistical tracking signal is considered to be robust for the following reasons.

In general a high tracking signal value may be caused by:

- a significant deviation in data pattern from the average estimate,
- an over-sensitive smoothing constant that yields over-responsive smoothed estimates, or
- large transients of random noise as data spikes.

For alerts from Toelle's crisp ETS there would still be doubt as to whether any transgressions were caused by an over-sensitive smoothing constant for data averaging, or whether they were caused by a single large transient. (We assume here that the initialisation tracking-in period had passed and that is not the cause.) On the other hand, the fuzzified ETS minimises such doubts because the use of TPD computations for the input granule, average, average error, and tracking signal itself, all serve to filter one-off transient input spikes by using only the most possible modal value of the FETS as the trigger for change detection. For these reasons, one has more confidence in attributing high FETS modal values to pattern changes in the data.

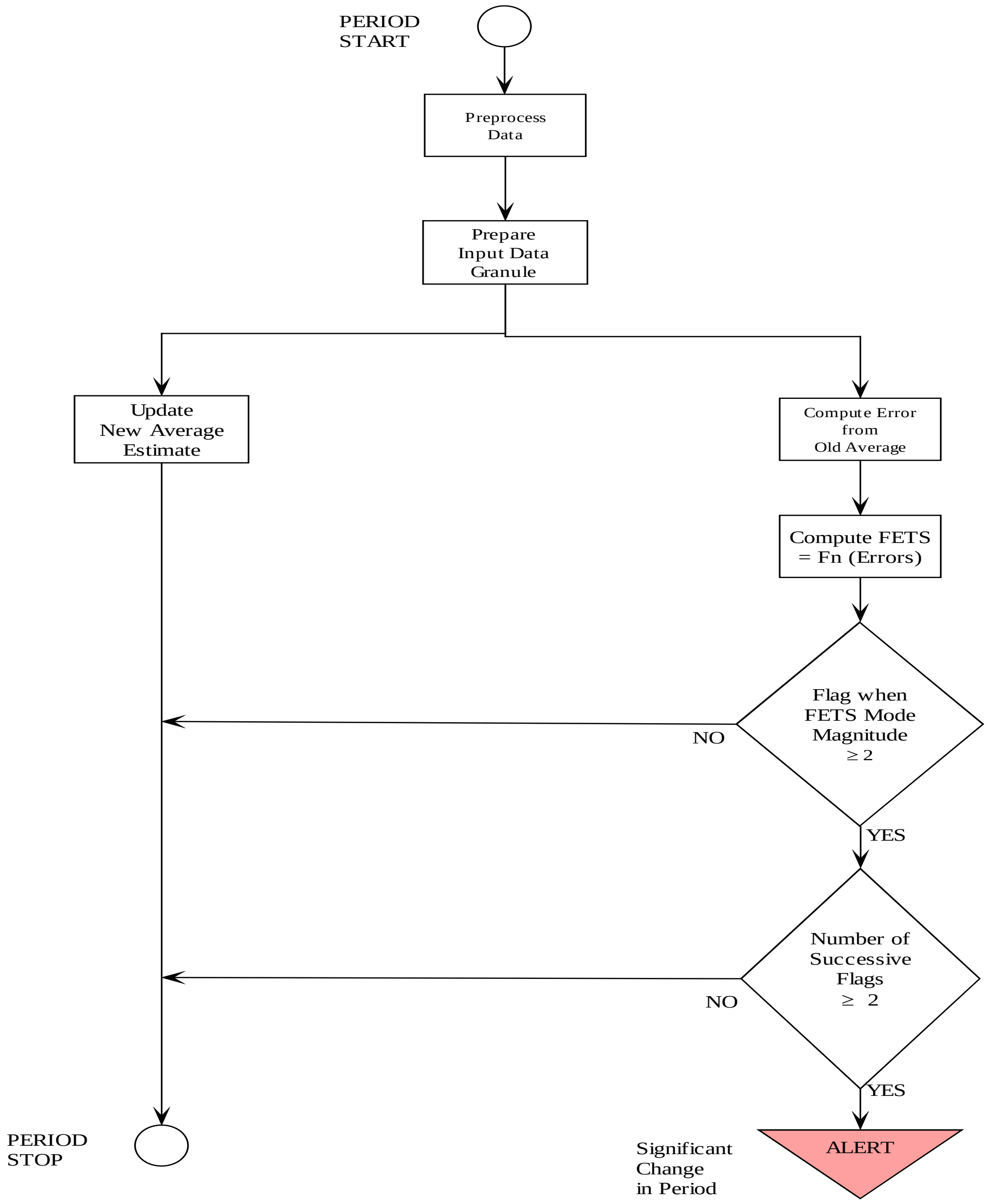


Figure 1: Main Computation Steps at Update Period

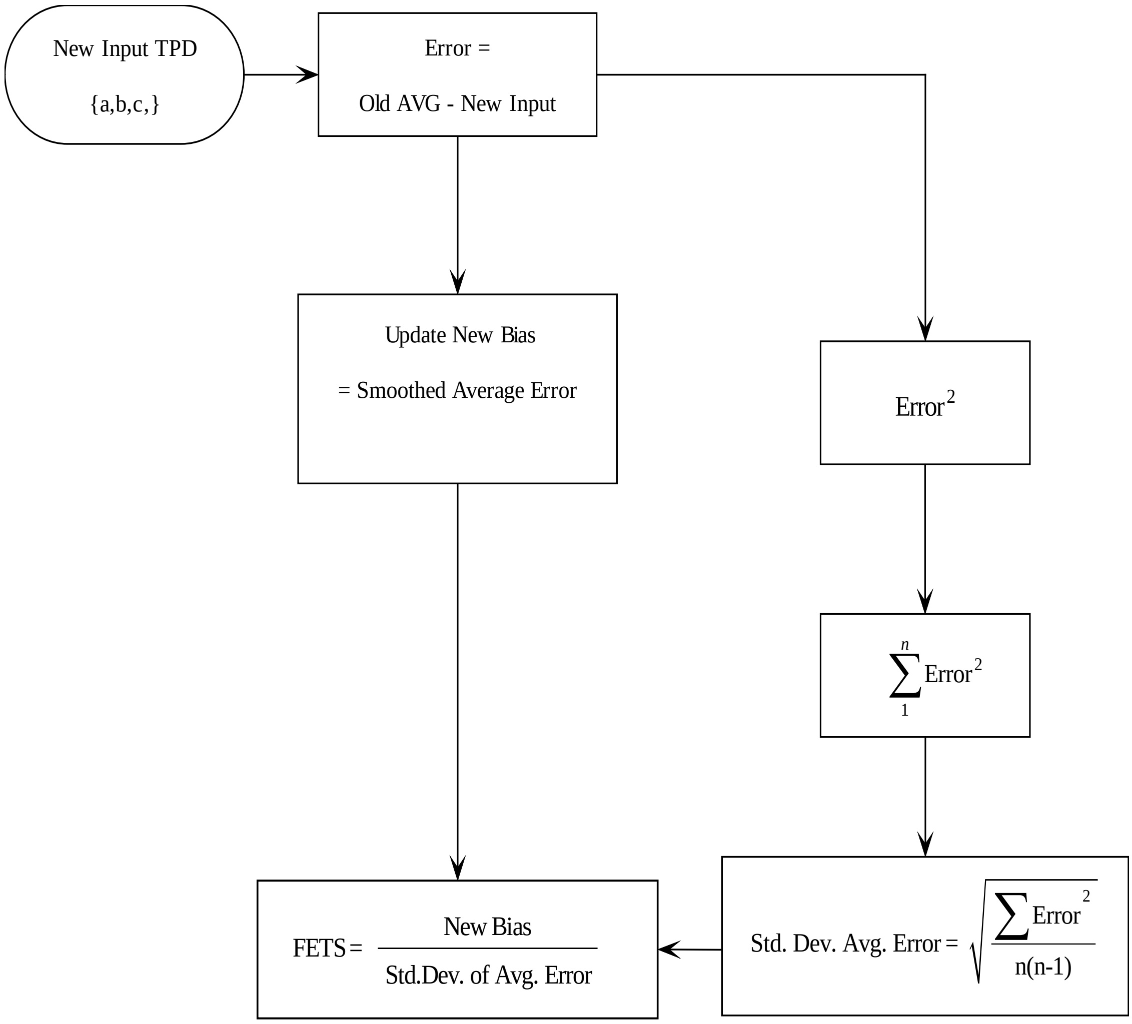


Figure 2: Fuzzy Enhanced Tracking Signal Computation Steps

4. Demonstration of Rough Time Series Tracking

The response behaviour of the FETS will now be demonstrated for three rough data series, with the computations for the first two applications shown in Tables 1 and 2.

4.1 Automated News Report Analysis

One application that can generate a rough series is the field of automated news report analysis as may be used for detecting or pre-empting crisis situations. The data in this example (Figure 3) is taken from the work by the Bonds [24] where they applied advanced parsing techniques to Reuters' news feeds. The subsequent derived measures shown in this data are based on Goldstein's Interstate Conflict evaluation methodology [25]. In this example, they represent the level of stress in the international relations of North Korea between beginning of 1990 and midway through 2000. A trend line shown in black indicates that the conflict level seems to be slowly decreasing.

Figure 4 illustrates the pre-processed data inputs and the mode of the exponential average granule, while Figure 5 shows the modal value of the FETS tracking signal.

The movement of the modal FETS demonstrates North Korea's well-known strategy of alternating threatening behaviour with cooperation gestures, most likely to maximise their power at the negotiating table in a brinkmanship manner. Negative modal values indicate increasing conflict levels while positive values indicate conciliatory periods. We will now correlate periods of high FETS modal values in Figure 5 with actual events that occurred.

From the input data (Figure 4) peaks in the pre-processed conflict measures exist around T = 18, 29, and 36.

The first FETS alert from T=17,18 (from 1st quarter '94) corresponds to a nuclear crisis that engulfed the peninsular around that time. The second FETS threat alerts around T=28-30 (from end '96) corresponds to a period of heightened tension caused by a North Korean infiltrator being killed and a North Korean submarine captured in South Korean waters. A subsequent smaller spike in the input pre-processed conflict measures at T=36,37 corresponds to North Korea's missile testing and the launch of a missile across Japan. However, this input spike did not generate a FETS alert because the input conflict measures were not sufficiently large and the high values only persisted for two periods. If greater sensitivity was desired to detect these smaller spikes, the value of the smoothing constant for updating the Bias could be increased (to 0.5 say from 0.4). In contrast to the negative value FETS alerts, FETS alerts of the opposite sign (+) correspond to significant conciliatory periods, at T=24 and 35-36, following the initial threat alerts (-) at T=18 and 29.

This example provides a degree of credibility to the pattern changes detected by the FETS, as well as demonstrating its robust nature. Notably, the FETS can provide more information than any overall trend line which can be misleading. Table 1 shows all TPD triplet computations for the input data set.

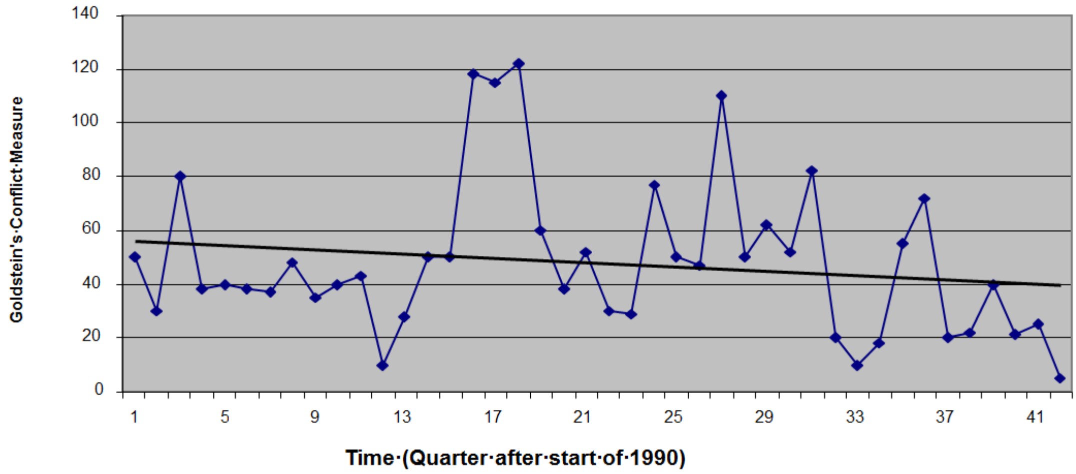


Figure 3: North Korea Interstate Conflict Measures

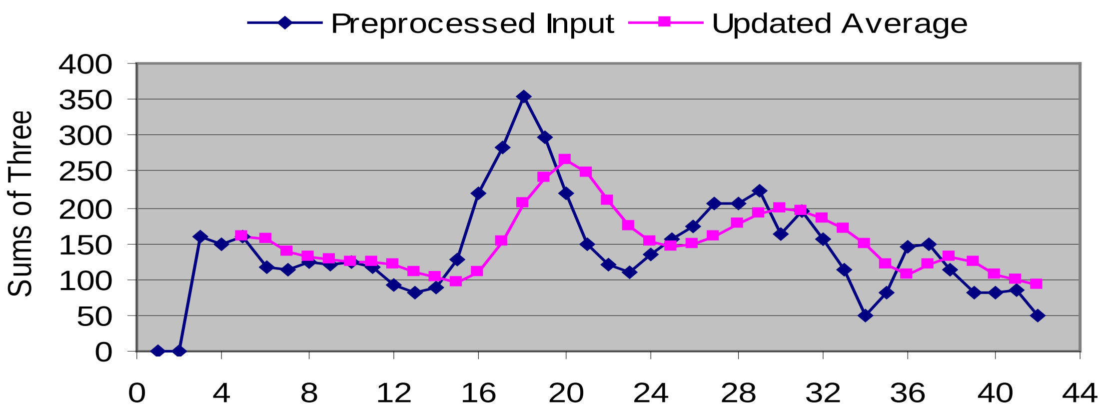


Figure 4: Pre-processed Input Data and Mode of Fuzzy Exponential Average

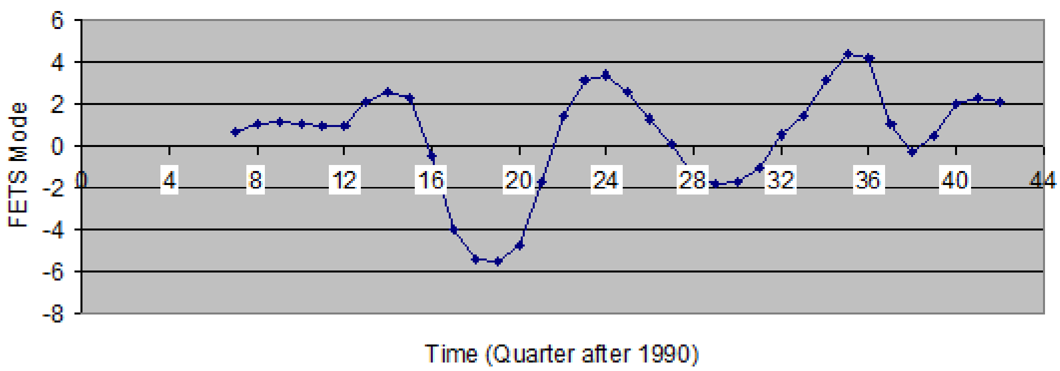


Figure 5: Fuzzy ETS Modal Value ('b')

Table 1: North Korea Threat Indicator

Period	Events	Events	Input TPD			Updated Ave			Error			Bias			Error^2			Sum Error^2			Variance			Std Deviation			ETS		
		Sum	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
1	50	0																											
2	30	0																											
3	80	160																											
4	38	148																											
5	40	158	148.00	158.00	160.00	148.00	158.00	160.00				-1.00	0.00	1.00															
6	38	116	116.00	148.00	158.00	135.20	154.00	159.20	-10.00	10.00	44.00	-4.60	4.00	18.20	100.00	100.00	1936.00	100.00	100.00	1936.00									
7	37	115	115.00	116.00	158.00	127.12	138.80	158.72	-22.80	38.00	44.20	-11.88	17.60	28.60	519.84	1444.00	1953.64	619.84	1544.00	3889.64	309.92	772.00	1944.82	17.60	27.78	44.10	-0.27	0.63	1.62
8	48	123	115.00	116.00	123.00	122.27	129.68	144.43	4.12	22.80	43.72	-5.48	19.68	34.65	16.97	519.84	1911.44	636.81	2063.84	5801.08	106.14	343.97	966.85	10.30	18.55	31.09	-0.18	1.06	3.36
9	35	120	115.00	120.00	123.00	119.36	125.81	135.86	-0.73	9.68	29.43	-3.58	15.68	32.56	0.53	93.70	866.24	637.34	2157.54	6667.32	53.11	179.80	555.61	7.29	13.41	23.57	-0.15	1.17	4.47
10	40	123	120.00	123.00	123.00	119.62	124.68	130.72	-3.64	2.81	15.86	-3.60	10.53	25.88	7.88	13.23	251.51	645.23	2170.77	6918.84	32.26	108.54	345.94	5.68	10.42	18.60	-0.19	1.01	4.56
11	43	118	118.00	120.00	123.00	118.97	122.81	127.63	-3.38	4.68	12.72	-3.51	8.19	20.61	11.44	21.95	161.68	656.67	2192.72	7080.52	21.89	73.09	236.02	4.68	8.55	15.36	-0.23	0.96	4.41
12	10	93	93.00	118.00	123.00	108.58	120.89	125.78	-4.03	4.81	34.63	-3.72	6.84	26.22	16.23	23.14	1199.19	672.90	2215.86	8279.71	16.02	52.76	197.14	4.00	7.26	14.04	-0.26	0.94	6.55
13	28	81	81.00	93.00	118.00	97.55	109.73	122.67	-9.42	27.89	44.78	-6.00	15.26	33.64	88.69	777.66	2005.03	761.59	2993.52	10284.74	13.60	53.46	183.66	3.69	7.31	13.55	-0.44	2.09	9.12
14	50	88	81.00	88.00	93.00	90.93	101.04	110.80	4.55	21.73	41.67	-1.78	17.85	36.85	20.70	472.28	1736.10	782.29	3465.80	12020.84	10.87	48.14	166.96	3.30	6.94	12.92	-0.14	2.57	11.18
15	50	128	81.00	88.00	128.00	86.96	95.82	117.68	-37.07	13.04	29.80	-15.90	15.92	34.03	170.02	888.04	1374.21	952.31	4353.83	13395.05	10.58	48.38	148.83	3.25	6.96	12.20	-1.30	2.29	10.46
16	118	218	88.00	128.00	218.00	87.37	108.69	157.81	-131.04	-32.18	29.68	-61.95	-3.32	32.29	880.90	1035.33	17172.06	1833.21	5389.16	30567.11	16.67	48.99	277.88	4.08	7.00	16.67	-3.72	-0.47	7.91
17	115	283	128.00	218.00	283.00	103.62	152.42	207.88	-195.63	-109.31	29.81	-115.42	-45.71	31.30	888.52	11947.78	38269.26	2721.73	17336.94	68836.37	20.62	131.34	521.49	4.54	11.46	22.84	-5.05	-3.99	6.89
18	122	355	218.00	283.00	355.00	149.37	204.65	266.73	-251.38	-130.58	-10.12	-169.80	-79.66	14.73	102.32	17052.06	63189.49	2824.04	34389.00	132025.86	18.10	220.44	846.32	4.25	14.85	29.09	-5.84	-5.37	3.46
19	60	297	283.00	297.00	355.00	202.82	241.59	302.04	-205.63	-92.35	-16.27	-184.13	-84.74	2.33	264.68	8528.55	42281.69	3088.73	42917.55	174307.54	16.97	235.81	957.73	4.12	15.36	30.95	-5.95	-5.52	0.57
20	38	220	220.00	297.00	355.00	209.69	263.75	323.22	-152.18	-55.41	82.04	-171.35	-73.01	34.21	3070.28	6730.32	23157.25	6159.00	49647.87	197464.79	29.33	236.42	940.31	5.42	15.38	30.66	-5.59	-4.75	6.32
21	52	150	150.00	220.00	297.00	185.82	246.25	312.73	-87.31	43.75	173.22	-137.73	-26.30	89.82	1914.41	7622.17	30006.25	8073.41	57270.04	227471.04	33.64	238.63	947.80	5.80	15.45	30.79	-4.47	-1.70	15.49
22	30	120	120.00	150.00	220.00	159.49	207.75	275.64	-34.18	96.25	192.73	-96.31	22.72	130.98	1168.48	9264.52	37146.34	9241.89	66534.56	264617.39	33.98	244.61	972.86	5.83	15.64	31.19	-3.09	1.45	22.47
23	29	111	111.00	120.00	150.00	140.09	172.65	225.38	9.49	87.75	164.64	-53.99	48.73	144.45	90.06	7700.31	27106.44	9331.96	74234.87	291723.82	30.50	242.60	953.35	5.52	15.58	30.88	-1.75	3.13	26.16
24	77	136	111.00	120.00	136.00	128.46	151.59	189.63	4.09	52.65	114.38	-30.76	50.30	132.42	16.76	2772.11	13083.74	9348.72	77006.98	304807.56	27.34	225.17	891.25	5.23	15.01	29.85	-1.03	3.35	25.33
25	50	156	111.00	136.00	156.00	121.47	145.35	176.18	-27.54	15.59	78.63	-29.47	36.42	110.91	243.06	758.65	6182.76	9591.78	77765.63	310990.32	25.24	204.65	818.40	5.02	14.31	28.61	-1.03	2.55	22.07
26	47	174	136.00	156.00	174.00	127.28	149.61	175.31	-52.53	-10.65	40.18	-38.69	17.59	82.61	113.33	1614.30	2758.99	9705.11	79379.92	313749.32	23.11	189.00	747.02	4.81	13.75	27.33	-1.42	1.28	17.19
27	110	207	156.00	174.00	207.00	138.77	159.37	187.98	-79.72	-24.39	19.31	-55.10	0.80	57.29	372.76	594.75	6354.59	10077.87	79974.67	320103.90	21.81	173.11	692.87	4.67	13.16	26.32	-2.09	0.06	12.27
28	50	207	174.00	207.00	207.00	152.86	178.42	195.59	-68.23	-47.63	13.98	-60.35	-18.57	39.97	195.56	2268.85	4655.25	10273.43	82243.52	324759.16	20.30	162.54	641.82	4.51	12.75	25.33	-2.38	-1.46	8.87
29	62	222	207.00	207.00	222.00	174.52	189.85	206.15	-69.14	-28.58	-11.41	-63.87	-22.58	19.42	130.18	816.79	4780.01	10403.61	83060.31	329539.17	18.85	150.47	596.99	4.34	12.27	24.43	-2.61	-1.84	4.47
30	52	164	164.00	207.00	222.00	170.31	196.71	212.49	-47.48	-17.15	42.15	-57.31	-20.40	28.51	294.04	1776.99	2254.60	10697.65	84837.29	331793.76	17.83	141.40	552.99	4.22	11.89	23.52	-2.44	-1.72	6.75
31	82	196	164.00	196.00	222.00	167.79	196.43	216.30	-51.69	0.71	48.49	-55.06	-11.96	36.50	0.51	2351.53	2671.81	10698.15	87188.82	334465.57	16.46	134.14	514.56	4.06	11.58	22.68	-2.43	-1.03	9.00
32	20	154	154.00	164.00	196.00	162.27	183.46	208.18	-28.21	32.43	62.30	-44.32	5.80	46.82	796.01	1051.50	3880.74	11494.17	88240.32	338346.31	16.37	125.70	481.97	4.05	11.21	21.95	-2.02	0.52	11.57
33	10	112	112.00	154.00	196.00	142.16	171.67	203.31	-33.73	29.46	96.18	-40.09	15.26	66.56	867.66	1137.59	9250.08	12361.83	89377.92	347596.39	16.35	118.22	459.78	4.04	10.87	21.44	-1.87	1.40	16.46
34	18	48	48.00	112.00	154.00	104.50	147.80	183.58	-11.84	59.67	155.31	-28.79	33.03	102.06	140.11	3560.95	24120.08	12501.94	92938.86	371716.47	15.40	114.46	457.78	3.92	10.70	21.40	-1.35	3.09	26.01
35	55	83	48.00	83.00	112.00	81.90	121.88	154.95	-7.50	64.80	135.58	-20.27	45.74	115.47	56.28	4199.58	18382.98	12558.23	97138.45	390099.44	14.43	111.65	448.39	3.80	10.57	21.18	-0.96	4.33	30.39
36	72	145	48.00	83.00	145.00	68.34	106.33	150.97	-63.10	38.88	106.95	-37.40	43.00	112.06	1511.85	3981.77	11438.37	14070.08	101120.22	401537.81	15.13	108.73	431.76	3.89	10.43	20.78	-1.80	4.12	28.81
37	20	147	83.00	145.00	147.00	74.20	121.80	149.38	-78.66	-38.67	67.97	-53.91	10.33	94.43	1495.41	4619.95	6187.52	15565.48	105740.17	407725.33	15.69	106.59	411.01	3.96	10.32	20.27	-2.66	1.00	23.84
38	22	114	114.00	145.00	147.00	90.12	131.08	148.43	-72.80	-23.20	35.38	-61.46	-3.08	70.81	538.35	1251.89	5299.33	16103.83	106992.06	413024.66	15.25	101.32	391.12	3.91	10.07	19.78	-3.11	-0.31	18.13
39	40	82	82.00	114.00	147.00	86.87	124.25	147.86	-56.88	17.08	66.43	-59.63	4.98	69.06	291.68	3235.09	4412.85	16395.51	110227.15	417437.50	14.61	98.24	372.05	3.82	9.91	19.29	-3.09	0.50	18.07
40	21	83	82.00	83.00	114.00	84.92	107.75	134.31	-27.13	41.25	65.86	-46.63	19.49	67.78	735.86	1701.33	4337.22	17131.37	111928.48	421774.72	14.40	94.06	354.43	3.79	9.70	18.83	-2.48	2.01	17.86
41	25	86	82.00	83.00	86.00	83.75	97.85	114.99	-1.08	24.75	52.31	-28.41	21.59	61.59	1.16	612.48	2736.81	17132.53	112540.96	424511.53	13.60	89.32	336.91	3.69	9.45	18.36	-1.55	2.28	16.70
42	5	51	51.00	83.00	86.00	70.65	91.91	103.39	-2.25	14.85	63.99	-17.94	18.89	62.55	5.04	220.49	4094.56	17137.57	112761.45	428606.09	12.87	84.66	321.78	3.59	9.20	17.94	-1.00	2.05	17.44

4.2 Tracking Insurgent Activity in Iraq

In 2007 Bowne [26] proposed a metric to help determine whether insurgent activity was increasing in Iraq. The metric focused on insurgent activity against coalition forces in pre-defined populated zones. Rather than measuring the actual number of casualties it simply records the number of *zones* in which one or more coalition death has occurred. Thus it is an indicator of the spread of insurgent activity. The claim was that such a measure is effective in identifying levels of insurgent behaviour because it is based on the regional expansion or contraction of insurgent activity.

The number of US monthly deaths in the Iraqi Freedom Campaign (as can be viewed at http://www.ac.wwu.edu/~stephan/USfatalities.html) since the start of the campaign in March 2003 up to March 2008 are: Year 1 (582), Year 2 (958), Year 3 (904), Year 4 (953) and Year 5 (455). Fitting a trend line (as at the above site) indicates a monthly death rate from 60 to 65 over the first five years. This low rate of change tends to indicate that Bowne's death zone count may be more useful for detecting change in insurgent behaviour. This seems possible because the actual number of deaths is dependent on many factors such as IED technology, delivery mechanisms, target size, plus many random circumstances.

Bowne also fits a trend line to his data (a red line in Figure 6) but it also demonstrates quite a shallow trend from about 11 to 16 zones per month over the shorter 3 year period of his data. Thus, Bowne concludes that there is a gradual increase in activity over the three years. Although some peaks can be identified in this graph it is not so clear at what points in time a significant shift (+ or -) has occurred. It would be very useful for military commanders to know if such a significant shift has occurred so they could relate that to any special efforts or strategies they might have implemented. To associate such a shift with any non-coalition causes would also be useful.

The mode of the smoothed three month sums average is shown in Figure 7, and Figure 8 shows the mode of the fuzzy ETS which tracks the sums as the data unfolds. Increasing negative FETS modal values indicate widening insurgent activities. The first point at which the magnitude of the mode is around 2 is at T=15,16 which corresponds to the peak in the data at T=13 and 14. The change indicated at this point is a significant increase in insurgent activity. The next significant change is at T=27,28 which indicates another significant step-up in activity. The final significant change is at T=36,37 which indicates a significant drop in activity. Table 2 details all TPD triplet values computed in developing the fuzzy tracking signal for this dataset.

In this way, the FETS has identified points where there have been probable changes in insurgent behaviour, which may not be so easily detected by visual examination of the data alone. These points would then need to be correlated with any implemented changes in military strategy or tactics. Automated implementation of the FETS could also alleviate the responsibility associated with human subjective interpretation of the data, or at least, be able to support any visual analysis. Automated implementation of the FETS could also provide a degree of interpretative consistency which might not be present across a range of different human observers with varying degrees of optimism and pessimism, or cognitive skills.

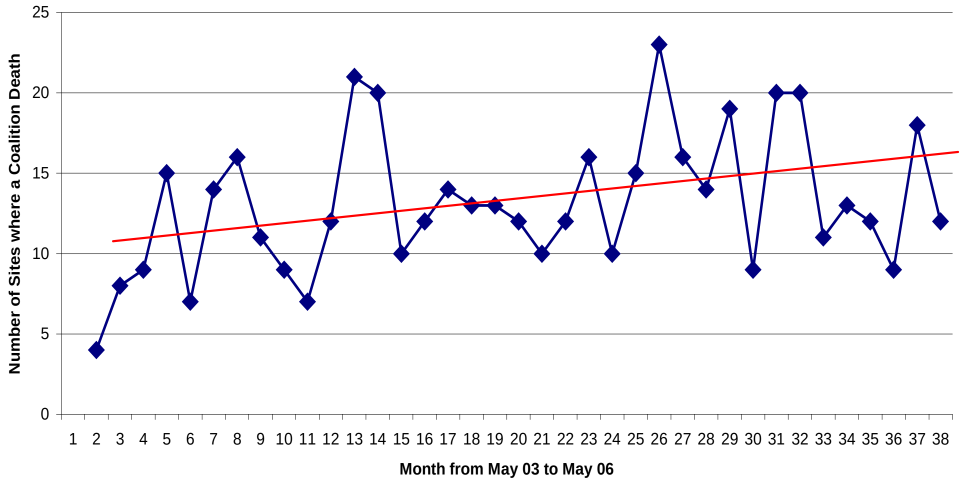


Figure 6: Number of Sites where a Coalition Death Occurred in a Month

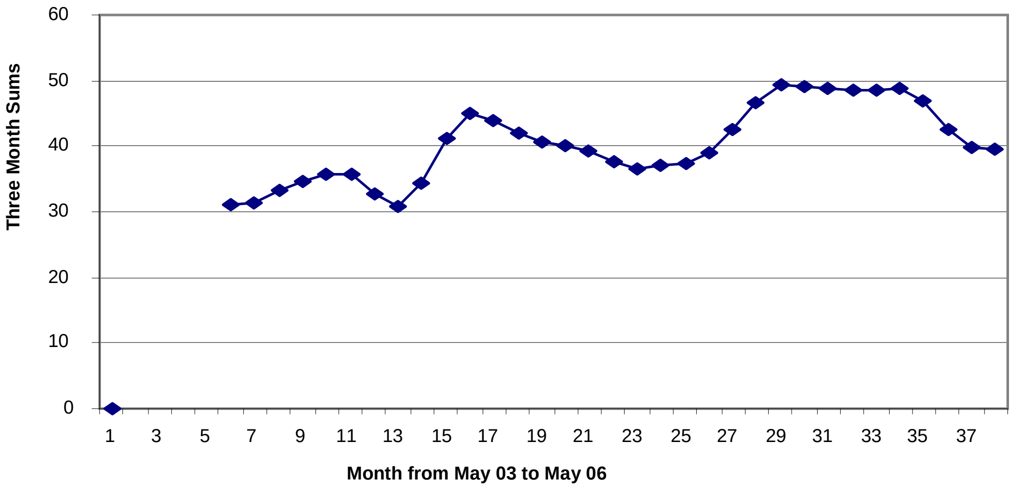


Figure 7: Mode of Smoothed Three Month Sums Average

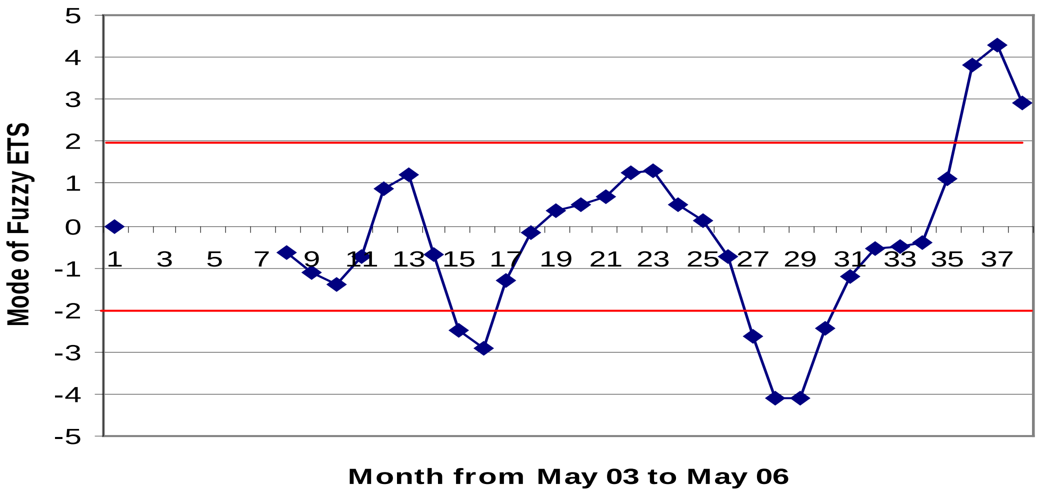


Figure 8: Mode of the Fuzzy ETS

Table 2: Iraq Insurgency Data Calculations

Period	Events	Events	Input TPD			Updated Ave			Error			Bias			Error^2			Sum Error^2			Variance			Std Deviation			ETS		
		Sum	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
1	4	0																											
2	8	0																											
3	9	21																											
4	15	32																											
5	7	31	21.00	31.00	32.00	21.00	31.00	32.00				-1.00	0.00	1.00															
6	14	36	31.00	32.00	36.00	25.00	31.40	33.60	-15.00	-1.00	1.00	-6.60	-0.40	1.00	1.00	1.00	225.00	1.00	1.00	225.00									
7	16	37	31.00	36.00	37.00	27.40	33.24	34.96	-12.00	-4.60	2.60	-8.76	-2.08	1.64	6.76	21.16	144.00	7.76	22.16	369.00	3.88	11.08	184.50	1.97	3.33	13.58	-0.64	-0.62	0.83
8	11	41	36.00	37.00	41.00	30.84	34.74	37.38	-13.60	-3.76	-1.04	-10.70	-2.75	0.57	1.08	14.14	184.96	8.84	36.30	553.96	1.47	6.05	92.33	1.21	2.46	9.61	-1.12	-1.11	0.47
9	9	36	36.00	37.00	41.00	32.90	35.65	38.83	-10.16	-2.26	1.38	-10.48	-2.55	0.89	1.89	5.09	103.23	10.73	41.39	657.19	0.89	3.45	54.77	0.95	1.86	7.40	-1.42	-1.38	0.94
10	7	27	27.00	36.00	41.00	30.54	35.79	39.70	-8.10	-0.35	11.83	-9.53	-1.67	5.26	0.13	65.55	139.84	10.86	106.93	797.03	0.54	5.35	39.85	0.74	2.31	6.31	-1.51	-0.72	7.14
11	12	28	27.00	28.00	36.00	29.13	32.67	38.22	-5.46	7.79	12.70	-7.90	2.11	8.24	29.79	60.65	161.17	40.65	167.58	958.20	1.35	5.59	31.94	1.16	2.36	5.65	-1.40	0.89	7.08
12	21	40	27.00	28.00	40.00	28.28	30.80	38.93	-10.87	4.67	11.22	-9.09	3.14	9.43	21.83	118.26	125.83	62.48	285.84	1084.03	1.49	6.81	25.81	1.22	2.61	5.08	-1.79	1.20	7.73
13	20	53	28.00	40.00	53.00	28.17	34.48	44.56	-24.72	-9.20	10.93	-15.34	-1.80	10.03	84.57	119.47	611.31	147.05	405.31	1695.34	2.63	7.24	30.27	1.62	2.69	5.50	-2.79	-0.67	6.19
14	10	51	40.00	51.00	53.00	32.90	41.09	47.93	-24.83	-16.52	4.56	-19.14	-7.69	7.84	20.78	272.84	616.77	167.83	678.15	2312.11	2.33	9.42	32.11	1.53	3.07	5.67	-3.38	-2.50	5.14
15	12	42	42.00	51.00	53.00	36.54	45.05	49.96	-20.10	-9.91	5.93	-19.52	-8.58	7.08	35.22	98.22	404.05	203.05	776.37	2716.16	2.26	8.63	30.18	1.50	2.94	5.49	-3.55	-2.92	4.71
16	14	36	36.00	42.00	51.00	36.32	43.83	50.38	-14.46	3.05	13.96	-17.50	-3.92	9.83	9.32	194.91	209.11	212.38	971.28	2925.26	1.93	8.83	26.59	1.39	2.97	5.16	-3.39	-1.32	7.08
17	13	39	36.00	39.00	42.00	36.19	41.90	47.03	-5.68	4.83	14.38	-12.77	-0.42	11.65	23.35	32.22	206.69	235.73	1003.50	3131.95	1.79	7.60	23.73	1.34	2.76	4.87	-2.62	-0.15	8.72
18	13	40	36.00	39.00	40.00	36.12	40.74	44.22	-3.81	2.90	11.03	-9.18	0.91	11.40	8.41	14.48	121.57	244.13	1017.98	3253.52	1.56	6.53	20.86	1.25	2.55	4.57	-2.01	0.35	9.11
19	12	38	38.00	39.00	40.00	36.87	40.04	42.53	-3.88	1.74	6.22	-7.06	1.24	9.33	3.03	15.08	38.63	247.16	1033.07	3292.15	1.36	5.68	18.09	1.17	2.38	4.25	-1.66	0.52	8.00
20	10	35	35.00	38.00	40.00	36.12	39.23	41.52	-3.13	2.04	7.53	-5.49	1.56	8.61	4.18	9.80	56.69	251.34	1042.86	3348.85	1.20	4.97	15.95	1.09	2.23	3.99	-1.37	0.70	7.87
21	12	34	34.00	35.00	38.00	35.27	37.54	40.11	-1.88	4.23	7.52	-4.05	2.63	8.17	3.53	17.86	56.51	254.86	1060.72	3405.36	1.06	4.42	14.19	1.03	2.10	3.77	-1.07	1.25	7.93
22	16	38	34.00	35.00	38.00	34.76	36.52	39.27	-2.73	2.54	6.11	-3.52	2.59	7.35	6.43	7.44	37.34	261.29	1068.16	3442.70	0.96	3.93	12.66	0.98	1.98	3.56	-0.99	1.31	7.50
23	10	38	34.00	38.00	38.00	34.46	37.11	38.76	-3.24	-1.48	5.27	-3.41	0.96	6.51	2.19	10.47	27.73	263.48	1078.63	3470.43	0.86	3.52	11.34	0.93	1.88	3.37	-1.01	0.51	7.02
24	15	41	38.00	38.00	41.00	35.88	37.47	39.66	-6.54	-0.89	0.76	-4.66	0.22	4.21	0.58	0.79	42.79	264.06	1079.42	3513.23	0.77	3.16	10.27	0.88	1.78	3.21	-1.45	0.13	4.79
25	23	48	38.00	41.00	48.00	36.73	38.88	42.99	-12.12	-3.53	1.66	-7.65	-1.28	3.19	2.74	12.48	147.02	266.80	1091.90	3660.24	0.70	2.87	9.63	0.84	1.70	3.10	-2.46	-0.75	3.81
26	16	54	41.00	48.00	54.00	38.44	42.53	47.40	-17.27	-9.12	1.99	-11.50	-4.42	2.71	3.97	83.16	298.43	270.77	1175.06	3958.67	0.64	2.80	9.43	0.80	1.67	3.07	-3.75	-2.64	3.38
27	14	53	48.00	53.00	54.00	42.26	46.72	50.04	-15.56	-10.47	-0.60	-13.12	-6.84	1.39	0.36	109.65	242.27	271.14	1284.71	4200.94	0.59	2.78	9.09	0.77	1.67	3.02	-4.35	-4.10	1.81
28	19	49	49.00	53.00	54.00	44.96	49.23	51.62	-11.74	-6.28	1.04	-12.57	-6.62	1.25	1.08	39.48	137.80	272.21	1324.19	4338.74	0.54	2.62	8.57	0.73	1.62	2.93	-4.29	-4.09	1.70
29	9	42	42.00	49.00	53.00	43.77	49.14	52.17	-8.04	0.23	9.62	-10.76	-3.88	4.60	0.05	64.70	92.59	272.27	1388.89	4431.34	0.49	2.52	8.03	0.70	1.59	2.83	-3.80	-2.44	6.55
30	20	48	42.00	48.00	49.00	43.06	48.68	50.90	-5.23	1.14	10.17	-8.55	-1.87	6.83	1.30	27.31	103.50	273.56	1416.20	4534.84	0.46	2.36	7.56	0.68	1.54	2.75	-3.11	-1.22	10.11
31	20	49	42.00	48.00	49.00	42.64	48.41	50.14	-5.94	0.68	8.90	-7.50	-0.85	7.66	0.47	35.23	79.28	274.03	1451.43	4614.12	0.42	2.23	7.10	0.65	1.49	2.66	-2.82	-0.57	11.79
32	11	51	48.00	49.00	51.00	44.78	48.65	50.49	-8.36	-0.59	2.14	-7.85	-0.75	5.45	0.35	4.59	69.91	274.38	1456.02	4684.03	0.39	2.07	6.67	0.63	1.44	2.58	-3.04	-0.52	8.72
33	13	44	44.00	49.00	51.00	44.47	48.79	50.69	-6.22	-0.35	6.49	-7.19	-0.59	5.87	0.13	38.65	42.06	274.50	1494.67	4726.09	0.36	1.98	6.25	0.60	1.41	2.50	-2.88	-0.42	9.73
34	12	36	36.00	44.00	51.00	41.08	46.87	50.81	-6.53	4.79	14.69	-6.93	1.56	9.40	22.92	42.64	215.83	297.42	1537.31	4941.93	0.37	1.89	6.09	0.61	1.38	2.47	-2.81	1.13	15.52
35	9	34	34.00	36.00	44.00	38.25	42.52	48.09	-2.92	10.87	16.81	-5.32	5.29	12.36	8.52	118.21	282.74	305.94	1655.52	5224.67	0.35	1.90	6.01	0.59	1.38	2.45	-2.17	3.83	20.85
36	18	39	34.00	36.00	39.00	36.55	39.91	44.45	-0.75	6.52	14.09	-3.49	5.78	13.05	0.56	42.56	198.50	306.50	1698.08	5423.16	0.33	1.83	5.83	0.57	1.35	2.41	-1.45	4.28	22.74
37	12	39	34.00	39.00	39.00	35.53	39.55	42.27	-2.45	0.91	10.45	-3.08	3.83	12.01	0.84	6.00	109.27	307.34	1704.08	5532.43	0.31	1.72	5.58	0.56	1.31	2.36	-1.30	2.93	21.58

4.3 Anomalies in the Murray-Darling Rainfall Data

4.3.1 Problem Definition

Recently, there has been a large public outcry in Australia about the current state of the Murray Darling Basin (MDB) river and lake system. There has also been widespread debate in recent times as to whether or not this is due to global climate change induced by man-made carbon emissions. Many experts are of the opinion that there is insufficient evidence to infer this as yet, and that it could be a long time into the future before we will be able to answer that with confidence. This author suggests that it is not necessary to make any assumptions about the causal effects of global climate change in order to establish a foundation upon which to base decisions about what corrective actions should be implemented for the MDB river system.

Rather, the primary question upon which corrective actions should be based should be:
Is the current drought part of a natural weather cycle which will largely be compensated for by heavy rain periods in the longer term, or is it an abnormally large anomaly (from whatever cause) which requires some prompt systemic intervention actions?

This section demonstrates that this question can be answered based on the results of applying the FETS to the MDB rainfall data.

4.3.2 The Murray Darling Basin Rainfall Data

The MDB annual rainfall data from the Australian Bureau of Meteorology (ABM) is shown below for the period 1900-2007, with a trend line added (Figure 9).

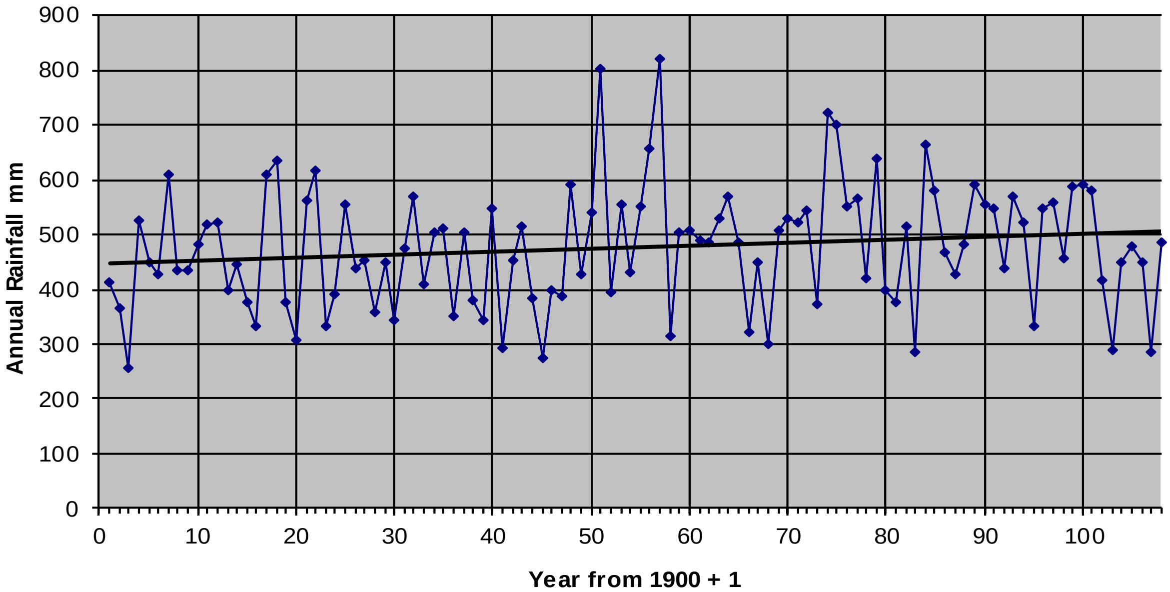


Figure 9: MDB Annual Rainfall from 1900 to 2007

4.3.3 Some simple methods for analysing time series

Trend analysis
Fitting a trend line as in Figure 9 indicates that over the whole period there is an ongoing increase in rainfall of about 0.60 mm per year. This result has been referred to by Jennifer Marohasy [27] of the Environment Unit in the Institute of Public Affairs, in her article in the Australian newspaper of August 23, 2008, "overall the trend is one of a slight increase in rainfall during the past 107 years."

A similar conclusion was also drawn in that article about a slight increase in rainfall over the period for the whole of Eastern Australia. However, the interpretation of scientific data is not always straightforward as Dr Marohasy states, "Many people want to save the environment but few people are confident of interpreting a chart or graph of scientific information on say, water quality or global temperatures."

However, one danger of making inferences based on long term averages, or trend, is that other important characteristics of the data are masked and neglected. The FETS method presented in this report will identify some other characteristics in the data that are very relevant to answering the initial question.

A Moving Average
Figure 10 shows the anomaly graph with an 11 year moving average shown in black as produced by the ABM. That moving average indicates that the period 1900-1950 suffered more periods of rainfall shortfalls than in the last 50 years, which indicates that the present drought period is nothing exceptional. However, in general a moving average lags behind recent data and is not very useful for identifying pattern anomalies.

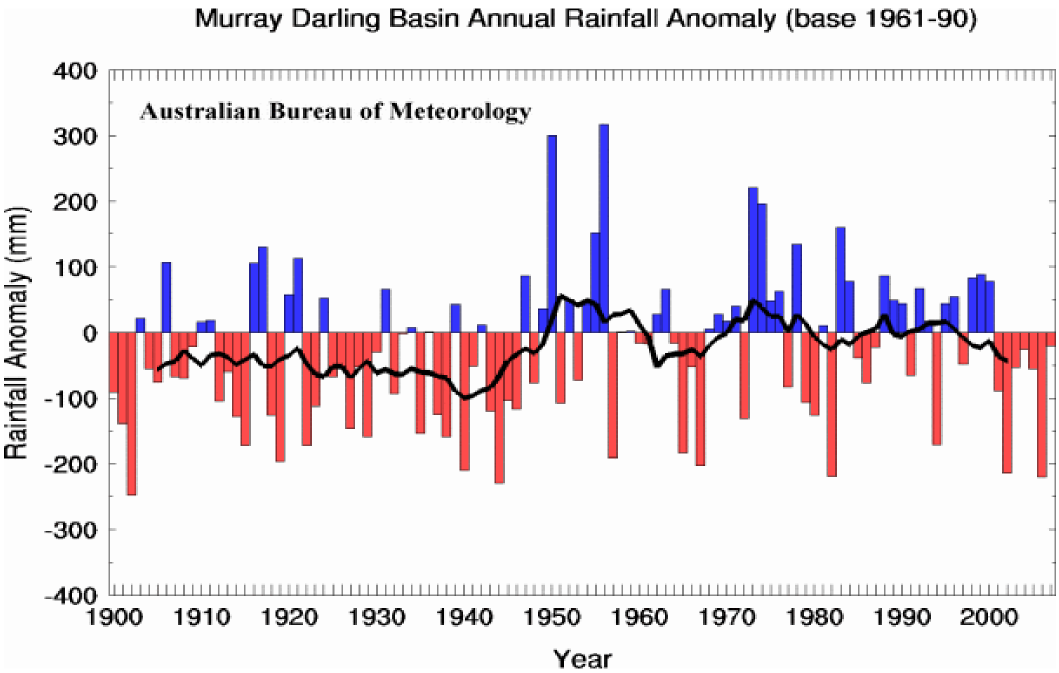


Figure 10: ABM's Anomaly Graph with an 11 Year Moving Average for the MDB

Figure 11 depicts the mode of the smoothed average of three year rainfall sums from 1900 to 2007. It can be noted that there is a step of about 200mm in the sum (i.e. about 66 mm per year) after the late forties, and a downward trend in the six years before 2007.

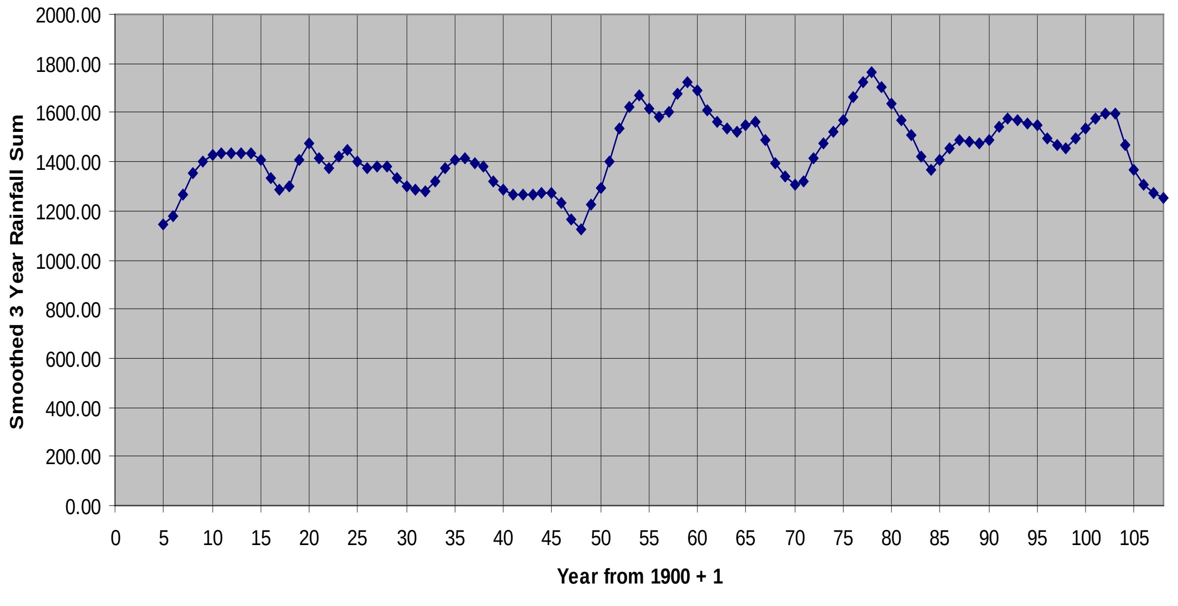


Figure 11: Smoothed Three Year Rainfall Sums (Mode)

Visual inspection

Across the data of recent years in Figure 9, we can see 7 successive years of low rainfall compared to the previous longest low run of 4 from 1943-1946. Beyond this simple information, visual inspection of the noisy series as it unfolds may also be used to identify some shifts in the rainfall pattern to a limited degree. For example, using a simple rule such as 3 or more years of low rainfall would cause serious depletion of the river system water storage levels, we could get an indication of when the MDB system is under stress. If this rule is applied to the unfolding data for MDB rainfall we would have raised drought or MDB stress alerts at the following years.

1902	Corresponds to the end of the so-called Federation drought that started in 1895.
1945-1946	Actual drought occurred (1940-1946)
2003-2007	Actual drought occurred.

So dynamic visual inspection using such a simple rule could identify some periods of drought.

4.3.4 Applying the FETS

The first step is not to base estimates of rainfall behaviour on rainfall in a single year, but rather to use granules of multiple years as the input data elements. Thus, the pre-processed input will be three year running sums as described previously. Figure 12 shows the response of the tracking signal (FETS) mode as each month of data is entered, after an initial tracking-in period of six years is allowed to initialise the computations.

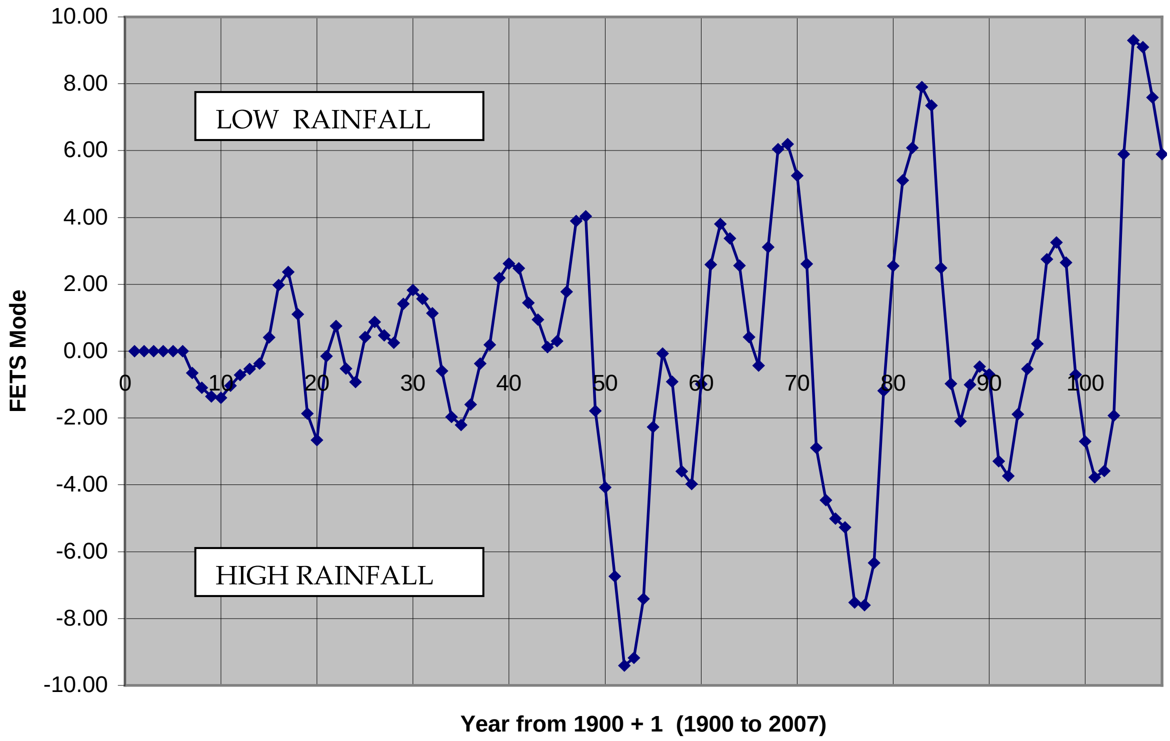


Figure 12: FETS Tracking Signal Mode for MDB Rainfall Data

When the FETS mode exceeds a magnitude of 2 twice in a row, a significant anomaly is indicated at the 95% confidence level. After two successive trips the FETS will continue generating alerts until the magnitude falls below 2. This behaviour will be demonstrated in the following summary tables. The positive side of the Y axis in Figure 12 indicates anomalies as a decrease in rainfall, while the negative side of the Y axis indicates anomalies in the direction of an increase in rainfall.

Both these anomalies are for three year sums of data and so are not based on single year spikes in annual rainfall. The high FETS values around 6-8 indicate that very large anomalies have occurred in both directions over the period. Since 1950 there seems to have been a cyclic pattern of anomalies in both directions with FETS peaks in high and low rainfalls showing an upward drift. What this means is that the cyclic occurrence of high and low rainfall figures is changing such that the periods of excess rainfall (which can cause floods) are becoming less extreme, while periods of low rainfall are becoming more extreme.

The commencement of the increase in cyclic amplitudes after 1950 also suggests that further research may be profitable on why that emerged over the last 60 years. The unique knowledge item provided by the FETS in Figure 12 is that the magnitude of the high and low rainfall periods is trending upwards towards the low rainfall side. In other words, even if the cyclic nature of high and low rainfall extremes does continue into the future, the periods of drought look like becoming more severe if the data pattern inherent in this

series continues. And although the fact that the pattern has occurred over the last 60 years does not necessarily mean that it will continue into the future, the rainfall in recent years in MDB does indicate that it is currently continuing. This conclusion about worsening droughts is then at odds with that drawn from the trend line in Figure 9 which indicates that the norm of MDB rainfall is gradually increasing (again with the proviso that such a trend does continue uninterrupted).

4.3.5 Correlation of FETS behaviour with significant MDB rainfall events

To further explore the knowledge provided by the FETS when it exceeds the control limit of 2 twice in succession, the dates of those alerts will be correlated with actual extreme events in the MDB to determine its accuracy. The method that will be applied is to update the FETS at the end of each calendar year based on moving 3 year sums of data. This means that the unit of time in the method is actually 3 years. An alert based on the most recent 3 years of data then indicates a significant change is currently in place. This change may then manifest an extreme event within the current time period (last 3 years including current year) or in the next time period (next 3 years). We will consider there to be a correspondence, or a hit, when an actual extreme event is in either of these two time periods as is shown in the tables below.

FETS Warnings of heavy rainfall or flood

The dates of actual floods shown apply specifically to Mildura, for which they have been validated. Notably, the weir system functioning since 1970s has also to a large extent controlled the river and minimised floods now in many locations. From Table 3 the total number of FETS flood alerts generated across the 108 year period is 15.

Table 3: Years that FETS Triggers Alerts of Heavy Rainfall Periods

Year of FETS Alert	1919	1934	1950-54 (5 times)	1972-77 (6 times)	1991	2000
Actual Floods (at Mildura)	None	1931	1952 1955 1956 (3 years)	1973 1974 (2 years)	1993	None

The FETS alerts corresponding to an actual event in a *current* period (last 3 years including current year) are at: 1952, 53, 54, 73, 74, 75, 76 (7 times)

The FETS alerts corresponding to an actual event in the *next* period (next 3 years from present) are at: 1950, 51, 52, 53, 72, 73, 91 (7 times)

From this data the alert years when corresponding events are in *either* period are:
1950, 51, 52, 53, 54, 72, 73, 74, 75, 76, 91 (11 times)

Thus, the proportion of flood alerts across the whole 108 year period = 15/108 = 14%
Proportion of flood alert hits for current periods = 7/15 = 47%
Proportion of flood alert hits for next periods = 7/15 = 47%
Proportion of flood alert hits in either period = 11/15 = 73%

Although the 2000 alert did follow three years of heavy rain, it did not seem to result in flood conditions anywhere. Also, the 1934 FETS warning occurred 4 years after 1931 so was not considered as a hit since it was beyond 3 years.

FETS Warnings of low rainfall periods or drought (< 400 mm p.a.)

Table 4 summarises the FETS low rainfall alerts with a total of 23 alerts.

Table 4: Years that FETS Triggers Alerts of Low Rainfall Periods

FETS Warning Year	1916	1939-1940	1946-1947	1961-1963	1967-1970	1980-84	1996-1997	2004-2007
Alert Period Range	1914-1919	1937-1943	1944-1950	1959-1966	1965-1973	1978-1987	1994-2000	2002-2010
Actual Low Rainfall Years in MDB (< 400mm)	1914 1915 1918 1919	1937 1938 1940 1943	1944 1945 1946		1965 1967 1972	1979 1980 1982	1994	2002 2006

The FETS alerts corresponding to a low rainfall event in a *current* period (last 3 years including most recent year) are at:
1916, 39, 40, 46, 47, 61, 67, 69, 80, 81, 82, 83, 96, 04, 06, 07 (16 times)
The FETS alerts corresponding to an event in the *next* period (next 3 years) are at:
1916, 39, 40, 69, 80, 04, 05 (7 times)
The FETS alert years when corresponding events are in *either* period are:
1916, 39, 40, 46, 47, 61, 67, 69, 80, 81, 82, 83, 96, 04, 05, 06, 07 (17 times)

Proportion of low rainfall alert years across whole period of 108 years = 23/108 = 21 %
Proportion of low rain alert hits for current periods = 16/23 = 70%
Proportion of low rain alert hits for next periods = 7/23 = 30%
Proportion of low rain alert hits in either period = 17/23 = 74%

4.3.6 Summary of FETS alerts for MDB rainfall data

The FETS graph in Figure 12 indicates that the rough rainfall pattern in the MDB data has been strongly cyclic with variable periodicity over the last 60 years, resulting in alternating periods of drought or very low rainfall, and flood or very high rainfall. Also from Figure 12, over the last 50 years the severity of the droughts has been increasing while the severity of the alternating high rainfall periods has been decreasing. This important inference must be considered in conjunction with the increase in the average rainfall over the last 50 years or so. Furthermore, the FETS alert success rate for both periods of extreme rainfall, or drought, is about 75% as determined above.

4.3.7 Conclusions about MDB rainfall data

The answer to the primary question stated in section 4.3.1 is that the drought at 2007 fits a rough cyclic rainfall pattern in the MDB that has strongly existed over the last 60 years, and more weakly in earlier periods. However, the drought from 2002 is an example of a persisting trend towards increasingly more severe droughts. This changing nature of expected droughts into the future thus indicates increasing stress on the MDB river and lake system such that pre-existing irrigation and river management strategies now need to be reviewed. And as previously stated, these conclusions require no assumptions about the existence of, or effects of, any global weather changes. Thus, even though the last 60years have shown an increase in the average rainfall over that of the first half of last century, at the same time there has been a trend across this period towards more severe droughts. Therefore, it would seem prudent to consider this increasing severity of droughts when planning MDB water management strategies, even though some experts state that the MDB rainfall data shows no unusual patterns, nor any trend to be concerned about over the longer term. It should also be noted that while the visual inspection rule suggested here could identify some droughts correctly, visual inspection could not derive the above conclusions. Finally, if we assume that the rainfall pattern in the data continued beyond 2007, and projected the cyclic FETS movement in Figure 12 beyond 2007, the possibility of heavier than average rain around 2012 is indicated. And this projection in fact corresponds to the heavy rain and floods that did occur in the MDB around 2010 to 2012.

5. Conclusions

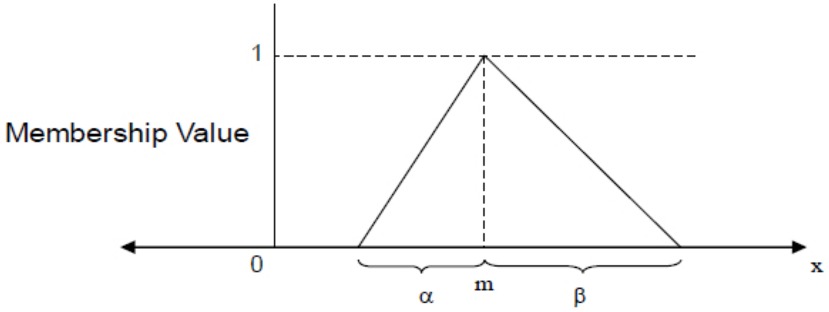
A hybrid fuzzy statistical tracking signal has been presented for promptly detecting when a significant pattern change or anomaly occurs in a rough sequence of data or qualitative information. This fuzzy tracking signal procedure essentially embeds a number of extensions to a tracking signal that was proposed by Trigg in 1964. The fuzzy tracking signal acts as a filter for short-term noise in such rough series, and helps to minimise unnecessary false alarms caused by non-enduring effects. Consequently, this tracking signal may facilitate the identification of anomalies in sequences of rough information as may be encountered in various Defence applications where time may be critical. A prototype tool to implement the fuzzy tracking signal procedure has also been developed.

6. References

[1] Zadeh, L. (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems*, **90**, pp.111-127.
[2] Martz, H. and Kvam, P. (1996) Detecting trends and patterns in reliability data over time using exponentially weighted moving averages, Reliability Engineering and System Safety, 51, pp. 201-207.
[3] Page, E. (1954) Continuous inspection schemes, Biometrika, 41, pp. 100-115.
[4] Spoerre, J. and Wang, B. (1995) Machine performance monitoring and fault classification using an exponentially weighted moving average scheme, Int. J. Production Research, 33(2), pp. 445-463.
[5] Tersine, R. (1994) Principles of Inventory and Materials Management, Prentice Hall: New York.
[6] Brown, R. (1959) Statistical Forecasting for Inventory Control, McGraw-Hill: New York.
[7] Brown, R. (1963) Smoothing, Forecasting and Prediction of Discrete Time Series, Prentice-Hall: New York.
[8] Gilchrist, W (1976) Statistical Forecasting, Wiley: New York.
[9] Dey, S. and Marcus, S. (1999) Change detection in Markov-modulated time series, Proc. Information, Decision and Control, Adelaide, pp. 21-24.
[10] Dey, S. and Jana, R. (2000) Change detection in teletraffic models, IEEE Trans. on Signal Processing, 48(3), pp. 846-853.
[11] Lee, W., Stolfo, S. and Mok, K. (1999) Mining in a data-flow environment: Experience in network intrusion detection, Proc. 1999 Knowledge Discovery in Databases Conference (KDD-99), San Diego, pp. 114-124.
[12] Fawcett, T. and Provost, F. (1997) Adaptive fraud detection, Data Mining and Knowledge Discovery, 1(3), pp. 291-316.
[13] Deboeck, G. (1994) Trading on the Edge, Wiley: New York.
[14] Ramirez, J., Cook, D., Peterson, L. and Peterson, D. (2000) An event set approach to sequence discovery in medical data, Intelligent Data Analysis, 4, pp. 513-530.
[15] Basseville, M. and Nikiforov, L. (1993) Detection of Abrupt Changes, Prentice-Hall: New York.
[16] Trigg, D. (1964) Monitoring a forecasting system, Operations Research Quarterly, 15, pp. 271-274.
[17] Deitrich, H. and Shafer, N, (1984) Business Statistics: An Inferential Approach, Dellen: San Francisco.
[18] Winston, W. (1997) Operations Research Applications and Algorithms, Duxbury Press: New York.
[19] Kaufmann, A. and Gupta, M. (1988) Fuzzy Mathematical Models in Engineering and Management Science, North Holland: Amsterdam.
[20] Hsoon, N. and Case, K. (1989) Development and evaluation of control charts using exponentially weighted moving averages, J. Quality Technology, 21, pp. 242-250.
[21] Toelle, R. (1993) Specification error in exponentially weighted moving average models, Technical Report, College of Business and Economics, Univ. Idaho.

[22] Toelle, R. (1994) Specification error in exponentially weighted moving average models, Proc. Annual Decision Science Institute Conference, Hawaii 1994, pp.1385-1387.
[23] Dubois, D. and Prade, H. (1980) Fuzzy Sets and Systems: Theory and Applications, Academic Press: New York.
[24] Bond, J. and Bond, D. (2000) Charting the North Korean Peninsula: A 1900-2000 Assessment of North Korea Event Interactions, Proc. 5th Int. Command and Control Research and Technology Symposium, October 2000, Canberra.
[25] Goldstein, J. (1992) A Conflict Cooperation Scale for WEIS Events Data, J. Conflict Resolution, 36(2) pp. 369-385.
[26] Bowne, D. (2007) Is the Insurgency Growing? A Novel Analysis of the Spatial and Temporal Dynamics of the Iraqi Insurgency, Defense and Security Analysis, 23(1) pp. 115-118.
[27] Marohasy, J. (2008) Case of the Warm and Fuzzy, The Australian Newspaper, August 23, 2008.

Appendix A: Fuzzy Triangular Numbers



Reference: Fuzzy Sets and Systems: Theory and Applications – Didier Dubois and Henri Prade, Academic Press, pp. 53-56, 1980

- The figure depicts a typical triangular fuzzy number (TFN). The key things to note are:
  - There is a single value of x (ie m) which has a membership value of 1
  - The TFN may be interpreted as representing a (triangular) possibility distribution for the vague concept of the "extent to which a real number is m"
  - The TFN is completely characterised by m and the two parameters $\alpha \geq 0$ and $\beta \geq 0$ which are referred to as the *left spread* and *right spread* respectively. Accordingly, the TFN can be represented by the triple (m, α, β)
- Using Zadeh's extension principle, ordinary arithmetic operations of addition, subtraction, multiplication (including scalar multiplication) and division can be extended to TFNs.
  - For addition, subtraction and scalar multiplication, the resulting fuzzy number is again a TFN but for multiplication and division the resulting fuzzy number is not a TFN, so it is customary to apply an approximation for these two operations. For two TFNs (m, α, β) and (n, γ, δ), the operations are as follows:
    - (Addition) $(m, \alpha, \beta) \oplus (n, \gamma, \delta) = (m + n, \alpha + \gamma, \beta + \delta)$
    - (Subtraction) $(m, \alpha, \beta) \varnothing (n, \gamma, \delta) = (m - n, \alpha + \delta, \beta + \gamma)$
    - (Scalar Multiplication) For $\lambda > 0$, $\lambda(m, \alpha, \beta) = (\lambda m, \lambda\alpha, \lambda\beta)$, while for $\lambda < 0$, $\lambda(m, \alpha, \beta) = (\lambda m, -\lambda\beta, -\lambda\alpha)$
    - (Multiplication) For $m, n \geq 0$, $(m, \alpha, \beta) \otimes (n, \gamma, \delta) = (mn, m\gamma + n\alpha, m\delta + n\beta)$
    - (Division) For $m \geq 0$, $n > 0$, $(m, \alpha, \beta) / (n, \gamma, \delta) = (m/n, (m\delta + n\alpha)/n^2, (m\gamma + n\beta)/n^2)$
  - Similar definitions of multiplication and division exist when one or both of m, n are negative – for details refer to pages 55 and 56 of the reference
  - In the special case that α and $\beta = 0$, the TFN reduces to a crisp number and in this case multiplication and scalar multiplication yield the same TFN

Page classification: UNCLASSIFIED

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION DOCUMENT CONTROL DATA	1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)
2. TITLE Change Detection in Rough Time Series	3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION) Document (U) Title (U) Abstract (U)
4. AUTHOR(S) Lewis Warren	5. CORPORATE AUTHOR DSTO Defence Science and Technology Organisation PO Box 1500 Edinburgh South Australia 5111 Australia

6a. DSTO NUMBER	6b. AR NUMBER	6c. TYPE OF REPORT	7. DOCUMENT DATE
DSTO-TR-3023	AR-016-091	Technical Report	June 2014

8. FILE NUMBER	9. TASK NUMBER		11. NO. OF PAGES	12. NO. OF REFERENCES
2014/1055711/1	NS 07/418	Head Language Technologies. and Fusion	25	27

13. URL on the World Wide Web	14. RELEASE AUTHORITY
http://dspace.dsto.defence.gov.au/dspace/	Chief, National Security and ISR Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CITATION IN OTHER DOCUMENTS Yes

18. DSTO RESEARCH LIBRARY THESAURUS http://web-vic.dsto.defence.gov.au/workareas/library/resources/dsto_thesaurus.shtml

Time series, Tracking algorithms, Fuzzy statistics

19. ABSTRACT

A discrete time series may characteristically have high noise levels resulting in a rough or jagged distribution which can present significant challenges to conventional statistical tracking techniques. To address this problem the proposed method applies hybrid fuzzy statistical techniques to series granules instead of to individual measures. After detailing the method and its rationale, three examples demonstrate the robust nature of the proposed fuzzy tracking signal which leads to a minimal number of false alarms caused by isolated spikes. The examples demonstrate the effectiveness of this tracking signal for promptly identifying significant pattern changes in rough time series as can be encountered in data sets used for various types of Defence decision making.

Page classification: UNCLASSIFIED